Journal of Economic Methodology, 2014
Vol. 21, No. 1, 54–76, http://dx.doi.org/10.1080/1350178X.2014.886473

Routledge
Taylor & Francis Group

Causal models and evidential pluralism in econometrics

Alessio Moneta[a]* and Federica Russo[b]

[a]*Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy;* [b]*Dipartimento di Studi Umanistici, Università di Ferrara, Italy*

(*Received 15 December 2011; accepted 23 October 2013*)

Social research, from economics to demography and epidemiology, makes extensive use of statistical models in order to establish causal relations. The question arises as to what guarantees the causal interpretation of such models. In this paper we focus on econometrics and advance the view that causal models are 'augmented' statistical models that incorporate important causal information which contributes to their causal interpretation. The primary objective of this paper is to argue that causal claims are established on the basis of a plurality of evidence. We discuss the consequences of 'evidential pluralism' in the context of econometric modelling.

Keywords: statistical model; causal model; evidential pluralism; causality; role of evidence

1. Introduction

Statistical models are widely used in quantitative causal analysis in the social sciences. A panoply of disciplines make use of them – from econometrics to sociology, from epidemiology to experimental psychology. 'Causal modelling', as it is also customarily called, has a long tradition tracing back to Quetelet (1869) and Durkheim [1897] (1960) who, in the nineteenth century, set the ground for a quantitative and scientific study of societal problems. Major improvements in quantitative causal analysis are also due to Wright (1921) and Haavelmo (1944) in the first half of the twentieth century and Blalock (1961) and Duncan (1975) in the second half of the twentieth century. In very recent noteworthy progress has been made by Pearl (2000) (and collaborators), Spirtes, Glymour, and Scheines (2000) and econometricians in the schools of Heckman (2008) and Hoover (2001), for example.

What distinguishes early methodologists from present-day ones (with a notable exception, perhaps, for Pearl (2000), Spirtes et al. (2000)) is that they were explicit in adopting a causal interpretation of the statistical models they developed. For instance, Wright had no doubt that his 'path coefficients' were measures of causal strength between two variables. Today's methodologists and practising scientists, on the contrary, are much more cautious in adopting a causal stance. This is, partly and possibly, because today's statistical models are much more complex, and also because it is certainly a very difficult task to establish stable causal relations in the unstable social world.

Although the caution (and sometimes scepticism) towards an overt causalist stance is understandable, the question remains as to what allows one to interpret, under circumstances to be specified, a statistical model in causal terms, no matter how complex it is. We address exactly this issue.

*Corresponding author. Email: amoneta@sssup.it

The paper draws on examples from causal modelling in econometrics and is organised as follows. In Section 2 we present statistical models used to infer associations and causal relations. We compare 'associational' and 'causal' models pointing at differences at three different levels: (1) background knowledge, (2) assumptions and (3) methodology. The presentation in this section draws upon the distinction between associational models and causal models, borrowed from Russo (2009a, 2009b), on the one hand, and between statistical and causal information, borrowed from Moneta (2007), on the other hand. In Section 3 we focus more specifically on the distinction between statistical and causal information, which is also akin to *statistical* versus *substantive* information (Spanos, 2006) and *statistical* versus *theory* information (Spanos, 1999). We argue that statistical models can be causally interpreted to the extent that statistical and causal information are carefully distinguished and their role assessed within associational and causal models, respectively. The idea of carefully distinguishing statistical and causal models was already discussed in Spanos (2005, 2010) and Hoover (2009), although in quite different terms (cf. also Hoover, Johansen, & Juselius, 2008). Our account introduces the notion of associational model and the idea of causal model as its *augmentation*. We argue that this augmentation is supported by a plurality of evidence. In Section 4 we discuss at length this thesis, which we refer to as 'evidential pluralism', and its consequences for econometric modelling. Throughout the paper, we highlight how evidence of mechanisms enters the stage of associational model and of causal model.

2. Associational models versus causal models

Scientists and philosophers alike will promptly agree on a broad distinction between associational models, on the one hand, and causal models, on the other hand. Here, we take 'associational models' as a handy short cut for 'statistical models in which associations among random variables are analysed', and 'causal models' as a handy short cut for 'augmented statistical models in which some of the associations can be interpreted as causal relations'. First, we clarify the distinction between associational and causal models, and then we present them in turn, focusing on three specific aspects: background knowledge, assumptions and methodology.

2.1 The distinction

Simply put, the claims made on the basis of an *associational model* refer to some properties or specifications of the probability distribution (joint or conditional) of the random variables that are studied. *Causal models* make a step further as they aim at uncovering *causal* relations. It is worth noting that associational models go beyond the so-called 'descriptive' statistics and make use of 'inferential statistics' as well: they aim to establish whether the data are a representative sample of a theoretical distribution (or data generating process [DGP]) and make inferences about the properties of this distribution from the data.

Notwithstanding the large agreement on this issue, more needs to be said. What is at stake is in fact to specify the features of associational models and of causal models in order to understand how causal models, under certain conditions, can use the information provided by associational models, such as measures of statistical dependence (e.g. regression coefficients or partial correlations), to support causal claims. We will also address this issue in Section 4 when discussing the evidence that supports causal claims. Notably, we will discuss how evidence of difference-making and evidence of mechanisms enter the two stages (associational and causal model) and support their respective inferences (about dependencies and about causal relations).

Table 1. Associational models versus causal models.

	Associational models	Causal models
Background knowledge	Choice of variables	Causal context; theoretical knowledge; institutional knowledge, etc.
Assumptions	Statistical	Statistical; extra-statistical; causal
Methodology	Model-based statistical induction and hypothetico-deductive method	Model-based induction and hypothetico-deductive method

In this section, we present, in order, associational models and causal models, pointing at differences at three levels: (1) background knowledge, (2) assumptions and (3) methodology. The difference between associational and causal models can be schematically represented as in Table 1.

2.2 Associational models

Associational models aim to provide an accurate and reliable description of how certain phenomena (chance events) are regularly associated among each other in their occurrence. Since the occurrence of chance events is quantitatively analysed through the lenses of the probabilistic structure of random variables, from an adequately specified associational model one can get a description of how the variations within the realisations of a random variable are (regularly) associated with the variations within the realisations of other random variables.[1]

In associational models, we start by analysing the data and infer a probabilistic structure underlying the observations. It is in this probabilistic structure that statistical dependencies between variables of interest, which are potentially meaningful, can be identified. Establishing statistical dependencies corresponds to what we will call in Section 4 'evidence of difference-making'. The causal interpretation is provided at a further stage of analysis (see discussion of causal models below), and the 'augmentation' also includes evidence of mechanisms, in a way that we discuss in Section 4. Moreover, dependencies are potentially meaningful, i.e. *not irrelevant*, only if the underlying statistically model is adequately specified. In the common statistical parlance, a measured probabilistic dependence between two variables in which there is no direct causal path connecting them is called *spurious*. This may due either to a third variable causing both of them (cf. common cause principle) or due to the mere fact that the probabilistic dependence stems from a mis-specified statistical model, which relies, for instance, on a too small data-set or on assumptions not adequately corroborated, so that the measured dependence is not a property of the underlying chance set-up (more on this below). It is important, however, to carefully distinguish between these two cases. In the case of misspecification, the alleged association would be irrelevant, whereas in the case of a common cause the statistical dependence would be 'spurious' but not irrelevant, since it delivers useful information. We argue that in an adequately specified statistical model, all the correctly identified statistical dependencies are *relevant*. Whether a statistical model is adequately specified is an empirical issue to be addressed with misspecification testing vis-à-vis the data, as suggested by Spanos (1999).

Associational models are formalised under the language of statistics so that associational models reduce to *statistical models*. Associational or statistical models, as

summarised in Table 1, are based on some background knowledge, assumptions (which includes model specifications) and methodology.

2.2.1 Background knowledge

Extra-statistical substantive knowledge plays a minor role in building an associational model. Since an associational model is built with the aim of bringing about evidence for theoretical questions, such questions influence the choice of the data of interest and the focus on particular dependence relations. However, once we choose the data and specify the focus of the model, background 'extra-statistical' knowledge does not play a crucial role any longer. This does not mean, of course, that in building an associational model the researcher needs to disregard any contextual (extra-statistical) knowledge. In fact, background knowledge may suggest, for example, that some macroeconomic data are likely to be difference-stationary or other consumption data regarding demographically different households non-homogeneous. These may actually be useful pieces of information. The point is that, in all these cases, the specifications of the associational model are based on the data *via* statistical tests.

2.2.2 Assumptions and specifications

Associational models are specified statistical models in which it is possible to analyse and estimate various measures and forms of statistical dependence. A (not yet specified) statistical model is a family of distribution functions describing a chance set-up (see Spanos, 1999). The most general way of describing a DGP (or chance set-up) is to say that each single observation is the realisation of a single random variable, where the relationships among the random variables are left open. The probability according to which a particular random variable assumes a certain value or a certain interval of values is determined by the probability distribution. Adopting the terminology of Spanos (2012), a particular data-set

$$\mathbf{Z}_0 = \{\mathbf{z}_{ik}; i = 1, \ldots, n, \; k = 1, \ldots, p\} \tag{1}$$

is viewed as the realisation of a set of random variables, i.e. of a *sample* $\mathbf{Z} = \{\mathbf{Z}_{ik}; i = 1, \ldots, n, \; k = 1, \ldots, p\}$. The statistical model determines the probabilistic structure of the sample. In its *general* form the statistical model can be written as:

$$\mathcal{M}_\theta(\mathbf{z}) = \{f(\mathbf{z}; \theta), \theta \in \Theta\} \;\; \mathbf{z} \in \mathbb{R}^{np}_{\mathbf{Z}}, \quad \text{for } \theta \in \Theta \subset \mathbb{R}^m, \;\; m < np \tag{2}$$

where $f(\mathbf{z}; \theta)$ denotes the joint distribution of the sample $\mathbf{Z}$ (cf. Spanos, 2011, 2012). This statistical model is so general that this is uninformative. Equation (2) just says that the data are generated by a chance set-up of which we do not know anything at all. Statistical inference precisely aims at obtaining a specified statistical model that formalises the chance set-up underlying the observed phenomena. Notice that, as Hacking (1965) has pointed out, the chance set-up can be merely hypothetical. What matters is that the phenomena display 'chance regularities patterns' (Spanos, 1999) such that the corresponding data can be described as generated by a chance set-up.

From a *specified* statistical model we can obtain useful statistical information for causal inference. The description of the chance set-up provided by a specified statistical model is much more informative than the description provided by the general statistical model. In fact, the specified statistical model partially opens the black box and provides some details about the chance set-up, while the general

statistical model just says that the data is generated by *some* chance set-up. The details of the chance set-up turn out to be useful to understand whether (and to what extent) variables are associated.

An important formulation of the statistical model in a manner which facilitates the further (data-based) specification and (causal knowledge-based) augmentation is obtained through the so-called 'marginal-conditional decomposition' (Mouchart & Russo, 2011). For example, let the data $\mathbf{Z}_0$ in Equation (1) correspond to $\mathbf{z}_{ik} = (x_i, y_i, w_i)$ for $i = 1, \ldots, n$ ($k = 3$). The marginal-conditional decomposition amounts to consider the set of random variables (X_i, Y_i, W_i) that have generated (x_i, y_i, w_i) for each i, to choose a set of conditioning variables (for example X_i and W_i) and to factorise the joint probability density function of (X_i, Y_i, W_i) in the following way:

$$f(X_i, Y_i, W_i; \theta_i) = f(X_i, W_i; \theta_i) f(Y_i | X_i, W_i; \theta_i), \quad \text{for } i = 1, \ldots, n \text{ and } \theta_i \in \Theta. \tag{3}$$

Needless to say, one can envisage alternative formulations by choosing alternative conditioning sets. Moreover, the (unconditional) joint density function on the right hand side of Equation (3) can be further decomposed into: $f(X_i | W_i; \theta_i) f(W_i; \theta_i)$.

Assuming that $E(|Y_i^2|) < \infty$, it is also possible to write:

$$Y_i = E(Y_i | X_i = x_i, W_i = w_i) + u_i, \quad \text{for } i = 1, \ldots, n \tag{4}$$

This is called by Spanos (1999, p. 370) a *statistical generating mechanism*, where $E(Y_i | X_i = x_i, W_i = w_i)$ is *the systematic component* and the term u_i is the *non-systematic component*, which is indeed obtained by subtracting the systematic component from Y_i.

The statistical model is further specified using statistical testing, which uses exclusively data and probability theory. For instance, suppose that after the appropriate testing, the random variables $(Y_1, \ldots, Y_n)$ turn out to be independent and identically distributed, and the same assumption can be made as regards both $(X_1, \ldots, X_n)$ and $(W_1, \ldots, W_n)$. Furthermore, suppose that the probability density function $f(X, Y, W)$ turns out to be normal.[2] In this case the statistical model (generating mechanism, in Spanos' terminology) reduces to the linear regression model:

$$Y = \alpha + \beta x + \gamma w + \varepsilon. \tag{5}$$

Background or causal knowledge can suggest further restrictions of this statistical model, which should not be in contrast with the testable specifications. For instance, substantive information may suggest that $\beta + \gamma = 1$ and it is possible to test this restriction through a standard F test.[3] If the restriction is consistent with the data, the statistical model (5) is reparametrised in $Y = \alpha + \beta x + (1 - \beta) w + \varepsilon$.

If the underlying distribution function is normal, we could use the partial correlation $\rho(X, Y | W)$ to measure the conditional dependence of X and Y given W. In particular, it turns out that, under the normality assumption, $\rho(X, Y | W) = 0$, if and only if X and Y are independent given W ($X \perp\!\!\!\perp Y | W$).[4]

If the statistical model is adequately specified, it accurately describes how variables co-vary. It is worth noting that the dependencies in the associational model are, to a certain extent, symmetric. This means that one may think of modelling $x = E(X | Y = y, W = w) + v$ instead of Equation (4) and obtain significant (although different) coefficient estimates. It will be the task of the causal modeller to add further constraints, typically coming from background knowledge, in order to attain the asymmetries sought for a

causal interpretation. As we shall see in the next section, the causal model will *augment* the associational model with a new viewpoint, which conceives of Equation (4) as stochastic mechanisms that generate y given x, w and u, provided that this direction of causal influence is compatible with background knowledge and other model assumptions. As mentioned above, alternative formulations of the statistical model are possible by choosing different (even opposite) conditioning sets.

2.2.3 Methodology

We have seen that in associational models we draw conclusions about statistical dependencies between variables from data. The 'machinery' of associational models helps establish that difference-making relations between variables hold (see Section 4). At a first sight, this is a pure inductive methodology: it uses a specific set of data to derive conclusions about the general stochastic process that generated the data. The inference is indeed inductive because it is both ampliative and not necessarily truth-preserving (it cannot be excluded that the conclusions inferred may be mistaken). These statistical dependencies are usually inferred from data within a hypothetico-deductive framework: we formulate hypotheses about the DGP and derive the consequences to be confronted with the data. The nature of these hypotheses is mainly statistical, while we will see that in causal models they have a more prominent causal import.

2.2.4 Example

Let us consider now a stylised econometric example of an associational model. Suppose we are interested in studying the relationship between expenditures on some good or service and income, and we have data sampled from a population of households (for example British households). Theoretical knowledge is crucial for the choice of variables and data, and particularly the choice of the level of aggregation of expenditures. The researcher can focus on categories of goods such as food, clothing and housing, or on more disaggregated items such as cereal products, milk, eggs, fat and milk. The choice of the level of aggregation usually depends on the research question one is interested in.

Available data (e.g. from the Family Expenditure Survey in UK) are usually based on survey of families, selected as a representative sample from a population (e.g. the population of British families), which is approximately considered as an infinite population. In this way the variations occurring in the underlying population can be formalised by a probability density function $f(X, Y; \theta)$, where Y denotes expenditure on some good and service, X denotes income and θ is a set of parameters. Such a statistical model presupposes already a specification, namely that the random variables $X_1, \ldots, X_n$ generating the household income data are independent and identically distributed, and the same can be said for the random variables $Y_1, \ldots, Y_n$ generating the household expenditure data.

In view of a causal analysis, the study of the regression model provides an important and useful specification of the statistical model:

$$m(x) = E(Y|X = x) = \int_{-\infty}^{+\infty} yf(Y|X)dy. \tag{6}$$

This regression function (conditional expectation) $m(x)$ can be specified and then estimated in a parametric way or alternatively directly estimated via a nonparametric method. For the sake of the estimation of this regression function, in the income–

expenditure setting referred to as *Engel curve*, it is not necessary to specify the functional form of the joint density $f(X, Y; \theta)$, although the marginal density function $f(X)$ in many empirical cases is significantly close to a log normal distribution. Examples of parametric specifications of Engel curves are:

$$Y = \log x + \varepsilon; \quad Y = x \log x + \varepsilon; \quad \log Y = \log x + \varepsilon; \quad Y = x^2 + \varepsilon \tag{7}$$

In the nonparametric approach we can write:

$$Y = m(x) + \varepsilon \tag{8}$$

where the function $m(x)$ is not specified a priori. In this case one finds, directly from the data, a set of points (whose connection constitutes a smooth line) which better fits the regression function. A common approach in empirical economics is to evaluate the adequacy of the parametric and nonparametric estimates of Engel curve by looking at measures of goodness of fit based on the smallness of the residuals. However, as Spanos (2007) points out, the adequacy of a statistical specification should not be based on goodness-of-fit measures, but rather on a criterion which is able to assess whether the specified statistical model accounts for the regularities in the data better than the alternative specifications. In the framework proposed by Spanos, which is in our view largely compatible with ours, an adequate specification of the model should be based on a test which is able to check whether the residuals are non-systematic enough (e.g. white noise). This test is empirical (based on observed data), once it is formulated in terms of a set of probabilistic assumptions about the DGP.

Notice that, in order to specify the regression function, extra-statistical background knowledge does play a role, but only at the beginning. Theoretical knowledge influences the choice of the data of interest and the expenditures Y on which the regression function curve is defined. But once the set of data (X, Y) is chosen, the estimation of the regression function curve is based only on statistical assumptions. These assumptions concern, in this specific case, the chosen functional form (in the parametric approach) and the method of estimation. The methodology is hypothetico-deductive, in the sense defined above: hypotheses about the functional form and distribution, which may, of course, be derived from background knowledge, are formalised and confronted with the data. The chosen 'direction' of the regression function (i.e. the estimation of $E(Y|X = x)$ instead of $E(X|Y = y)$) depends on the causal model we eventually want to analyse. Indeed, only if the underlying causal influence runs from X to Y and not from Y to X, the coefficient estimates of $m(x)$ will make substantive sense. But these considerations enter only in the specification of the corresponding causal model (see the next section).

2.3 Causal models

To go beyond claims about associations and statistical dependencies we need (1) more background (extra-statistical) knowledge of the causal context, (2) further assumptions and (3) a methodology to confirm/disconfirm causal hypotheses. These features will be here sketched and discussed more thoroughly in Section 3, where we will pay special attention to the 'extra-statistical' assumptions made in causal models.

2.3.1 Background knowledge

Background knowledge used in causal models includes general knowledge of the socio-political context and of the demographic characteristics of the population under analysis. It also includes economic theory or 'institutional knowledge' such as, for example, the functioning and procedures of a central bank. In disciplines at the border between the social and the biomedical, such as epidemiology, background knowledge also includes information about the physical-biological-physiological mechanisms. In Section 4 we will develop further this point, namely how evidence of mechanisms supports inferences about causal relations and how this is related to evidence of difference-making, established in the form of dependencies in associational models.

2.3.2 Assumptions

Causal models have the same 'statistical' assumptions described above for associational models. But, in addition, they also have 'extra-statistical' assumptions. We can classify under 'extra-statistical' assumptions all assumptions that are not related to the statistical properties of the distribution and that have causal meaning or that impose restrictions on the distribution but are not subject to statistical testing. For instance, the direction of time or the direction of the causal relation belong to this class. The causal Markov condition in the graphical models literature (Spirtes et al., 2000) is an example of an assumption imposing restrictions on the probability density function but not subject to direct statistical testing. There could be other assumptions, which are instead subject to statistical test and which can directly contribute to interpreting the relations between the variables in the model causally. Causal assumptions will be more thoroughly discussed in Section 3.

2.3.3 Methodology

In causal models a common way to draw causal conclusions is to put in place a hypothetico-deductive methodology. In very simple terms, we first formulate hypotheses out of background knowledge and of preliminary analyses of data and then we test them to ascertain whether and to what extent the consequences derived from such hypotheses hold. The possible causal structures, in order to bridge the gap with statistical models, are typically formulated as *structural models*, which are systems of equations which can be easily confronted with the system of equation obtained from the specified statistical model. The so-called *identification problem* (cf. Dufour & Hsiao, 2008) studies the conditions under which this confrontation is possible. If there are more estimated statistical parameters than unknown structural parameters, the system is under-identified; in the opposite case (more statistical than structural parameters), the system is over-identified. If the number of unknowns is equal to the number of knowns, the system is just-identified. Since only over-identified systems are statistical testable, it is typical that the researcher imposes zero restrictions on the parameters in order to get testable hypotheses.

A way to get over-identifying restrictions is through the graphical-model procedure proposed by Spirtes et al. (2000) and Pearl (2000). To be correctly performed, our perspective suggests that this procedure should be based on an adequately specified statistical model. The first step of this method, aimed at testing conditional independence relations, is consistent with the statistical-model methodology described above. In the second step, under certain conditions or rules of inference such as the causal Markov condition and the faithfulness condition, a causal graph representing the possible causal structures is derived from the conditional independence tests.

2.3.4 Example

Let us consider again the econometric example introduced in Section 2.2. Under what conditions can regression functions such as Engel curves be used to run counterfactual experiments? Suppose one is interested in evaluating the effect of a tax reform, which will affect household income and, indirectly, household expenditure patterns. Engel curves, as regression functions, are framed in associational models and therefore do not necessarily remain stable under changes of income, and may even change over time. When Engel curves are formalised in the framework of a causal model, then they can be used to evaluate policy intervention. This means that Engel curves are augmented with theoretical knowledge on consumption behaviour (extra-statistical assumptions). In Section 4 we characterise such theoretical knowledge as evidence of mechanisms; in the discussion of Engel curves, we show how information concerning socio-economic mechanisms justifies augmenting the associational model to a causal model.

A common approach is to formulate a 'demand system', namely a system of equations which explicitly formalises the effect of income and price on consumption (Banks, Blundell, & Lewbel, 1997; Deaton, 1986). These equations are based on theoretical assumptions about consumer behaviour, and the typical assumption in mainstream economics is that individuals maximise their utility function. In this manner the system is meant to remain stable under interventions that yield changes in income and prices. Since there are alternative specifications of the demand system (in terms, for example, of the different functional forms that describe the dependence of expenditure on income and prices), it is important to choose the one that better matches the consumption patterns, as they emerge in the study of the statistical Engel curves. The methodology is therefore hypothetico-deductive: one first formulates hypotheses out of economic theory (rational choice theory in mainstream economics) and then confronts them with the data.

It seems quite clear from the presentation above that associational models and causal models cannot reach the same goals because they have different *apparati*. We want to suggest that causal models could be thought of as *augmented* associational models, augmented exactly in some particular assumptions, in the amount of background knowledge used and in the model-building and model-testing methodology. These extra features are the extra tools that make causal inference possible in causal models but not in associational models. In the next section we examine in more details the import of statistical and causal information in this augmentation from associational to causal models.

3. Statistical and causal information

In this section we elaborate more on the meaning and import of statistical and causal information in associational and causal models. We will argue that while associational models only convey statistical information, causal models also convey *causal* information. We will specifically address two questions. What is the difference between statistical and causal information? And mostly, does 'causal information' introduce a vicious circle?

3.1 Statistical information

Statistical information, according to an established tradition, is a summary of data, or as Fisher (1922, p. 311) put it, a 'reduction of data':

> A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in

other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

This reduction is accomplished by constructing a model which formalises the (hypothetical) population or process from which the observed data are generated. This latter feature is what makes the collection of statistical information a procedure governed by inferential statistics, as distinguished from descriptive statistics. While descriptive statistics limits its analysis to the sampled data, inferential statistics goes beyond the sample and seeks to reach conclusions about the whole population. Thus, collecting statistical information, and in particular statistical information which is useful for causal inference, amounts to obtaining information about some features of the process which has generated the data and not just about the 'superficial' aspects of the data. Later in Section 4 we will advance the view that statistical information corresponds to what the recent literature on evidential pluralism has called 'evidence of difference-making'.

In this way, statistical information delivers an adequate and parsimonious description of phenomena. It is important to ask what kinds of phenomena, since not all the phenomena are object of statistical analysis. Statistical information is about phenomena which exhibit 'chance regularity patterns' (Spanos, 1999).[5] These are phenomena whose singular occurrence is uncertain, i.e. we do not know which particular value a variable is going to take, and, at the same time, phenomena whose aggregate occurrence has some order, i.e. we know that among many events a certain range of values will regularly come out.[6]

There is a particular piece of statistical information which is especially useful for causal inference. This is the notion of *statistical dependence*. Statistical dependence establishes that some events are frequently associated in their occurrence. Note that the definition of statistical dependence (and independence) merely involves the notion of density function:

$$X \text{ and } Y \text{ are dependent} \quad \text{iff } f_{XY}(x,y) \neq f_X(x)f_Y(y). \tag{9}$$

Conversely, using the symbol of statistical independence $\perp\!\!\!\perp$:

$$X \perp\!\!\!\perp Y \quad \text{iff } f_{XY}(x,y) = f_X(x)f_Y(y). \tag{10}$$

An important related notion is that of *conditional independence*. Conditional independence involves the notion of conditional density function and is used as input for many causal search algorithms (cf. Pearl, 2000; Spirtes et al., 2000):

$$X \perp\!\!\!\perp Y|W \quad \text{iff } f_{XY|W}(x,y|w) = f_{X|W}(x|w)f_{Y|W}(y|w) \tag{11}$$

It is important to emphasise that the notions of statistical dependence, independence and conditional independence are formalised within the framework of the statistical model: they are indeed characteristics of the chance set-up and only indirectly of the data. The data are interpreted through the lenses of the statistical model, which has the capacity of generating data. The observed data are indeed seen as a *typical* realisation of the statistical model. Thus, when we adequately identify a statistical-dependence relation, we have captured an important aspect of the chance set-up (two variables are regularly associated) and, therefore, of the phenomena of interest, which are realisations of this set-up. The model is specified in terms of probability density functions and then further specified in terms of a system of stochastic equations. Since dependence and the related notions involve only probability densities, it is apparent that they are *first* properties of the

model and *then*, if the model accurately describes the process generating the data, also of the data. We will come back to this idea of 'model-dependence' later in Section 4, but it is worth saying at this point that model-dependence is not a consequence of an anti-realist position. As Hoover (2009) also says:

> Without the models, there are no probabilities to discuss. This is not an anti-realist thesis.

To put it in another way, if we take any data and mechanically calculate correlations, without thinking of the underlying chance set-up, it is very likely that we are led astray. Consider a stock example discussed in the philosophical literature (Sober, 2001): if we calculate the correlation between sea levels in Venice and British bread prices and we do not pay attention to the characteristics of the underlying random variables and respective probability distributions, we may think to have identified statistical dependencies, but actually, in our perspective, we have just estimated properties of (most likely) inadequately specified statistical models (see Hoover, 2009, for a similar discussion).

It is worth noting that correlation is a measure of dependence, but the two notions do not necessarily coincide. In particular, if two variables are uncorrelated, this does imply that they are independent. However, there are special cases where this is true. For instance, if the probability distribution is Gaussian, the statistical model shows a bi-implication between zero correlation and independence.

Statistical information depends on two aspects: (1) the statistical model in which it is framed and (2) the specific hypotheses about the DGP made in the statistical model. The DGP can also be interpreted as a stochastic mechanism. However, it is important to notice that the statistical model *postulates* a stochastic mechanism, but does not describe it in full detail: it only assumes that the postulated mechanism is able to generate chance regularity phenomena, of which the observed data are a representative sample. This point of view of the statistical modeller is appropriately described by Lindley (2000, p. 295):

> [. . .] it is only the manipulation of uncertainty that interests us. We are not concerned with the matter that is uncertain. Thus we do not study the mechanism of rain; only whether it will rain.

3.2 Causal information

Causal information goes beyond mere association and in this sense opens the 'black box' of the DGP a step further by providing more details about such mechanisms. Causal information is 'augmented' statistical information: it allows additional interpretation so that an association between, say, two variables X and Y can be viewed as, for example, a causal influence from X to Y. But the causal influence may remain stochastic, either because we do not know the exact value of additional variables further influencing Y,[7] or because the influence of X on Y may be inherently probabilistic. The additional interpretation making statistical information *causal* comes from extra-statistical assumptions which, as we have seen, rely on different kinds of 'background knowledge', for instance theoretical knowledge, information about institutional mechanisms, or views on the nature of causality (e.g. temporal priority of cause) and its relations with the notion of statistical dependence.

Before addressing the issue about the conditions under which causal interpretation is guaranteed (see also next section), it is important to clarify the extent to which causal information is shaped by statistical information and background knowledge. Statistical information provides the formalised empirical evidence which can be used to test hypotheses about causal claims. In the next section we will suggest that statistical evidence

codifies evidence of difference-making, namely that there is an appropriate difference-making relation between the putative cause and the putative effect. However, not all hypotheses can be tested. Background knowledge provides the a priori constraints on this information: some causal relations are excluded or allowed on the basis of that.

For example, in studying the influences of consumption decisions of families, one may decide to consider the disposable income as exogenous to this decision, relying on occasional observations of families or economic theory of consumption behaviour. In this way, background knowledge a priori excludes (arbitrarily or not) that household expenditures do not cause household income and that the latter causes the former, which explains their statistical dependence. In other contexts, background knowledge may take a completely different form, and be formalised as 'rules of inference', i.e. assumptions on the relation between causal properties and conditional independence relations.

Suppose we know all the possible conditional independence relations between three random variables X, Y, and W, and in particular we know that X and Y are not statistically independent, that Y and W are not statistically independent, but that X is independent of W given Y, in symbols $X \perp\!\!\!\perp W|Y$. According to two rules of inference on which several graphical methods for causal inference are based,[8] since Y 'screens off' X from W, some causal structures are excluded and others are allowed. In graphical representations, the excluded structures are: $X \rightarrow Y \leftarrow W$ and any structure in which X and W are directly connected. The allowed structures are: $X \rightarrow Y \rightarrow W$, $X \leftarrow Y \leftarrow W$ or $X \leftarrow Y \rightarrow W$.

We now turn to the second question. Does the dependence of causal information on statistical information and background knowledge lead to circularity? One could in fact argue that statistical information and background knowledge depend, in turn, on causal information. Statistical information, being framed within a statistical model, depends on the postulation of a DGP. We have seen that, at the stage of building a statistical model, extra-statistical background information may play some role, but this is non-causal knowledge, and even if it is causal, it does not relate to the possible causal relationships among the objects under scrutiny. Moreover, the mechanisms operating in DGP have not actually been uncovered: what counts are the concomitant variations of the events. Thus, statistical information in an associational model is actually *not* dependent on causal information.

On the contrary, background knowledge does depend, to a certain extent, on causal information and so the peril of having hidden causal assumptions (about the sought causal relations) in the background knowledge used is real. This circularity is avoided by carefully choosing the 'parts' of background knowledge that we are entitled to use. Specific assumptions about the causal relations under study should be explicitly excluded. For example, if we are trying to understand whether some household expenditures may influence future household income (some expenditures on housing and durable goods may be seen as partial forms of investment and generate future income), we should not build a statistical model in which income is an exogenous variable.

But this triggers another important and related question: how much background knowledge is really needed to get *new* causal information? In the methodology of applied research, such as econometrics, this is quite a debated question. In the literature, different answers have been given, and they paved the way to different traditions (see Moneta, 2007). In particular, 'inductivist' approaches seek to minimise the use of background knowledge, while 'deductivist' approaches eschew statistical information as a tool of causal discovery and rely only on theoretical knowledge for the articulation of a causal structure (i.e. economic theory provides all the background knowledge needed). These two

approaches, may, quite clearly, lead to two tendencies: inductivist approaches are at risk of not being able to make the 'big leap' beyond statistical information, while deductivist approaches run the risk of being stuck in a vicious circle.

Our answer about the amount of background knowledge we need is: *it depends*. There are situations where we have at our disposal a large volume of data, which may allow us to infer a very detailed picture of the statistical dependencies (through conditional independence tests); also, in case we can safely assume that the DGP is stable, only a relatively limited set of extra assumptions and rules of inference may suffice (as those proposed by Pearl (2000); Spirtes et al. (2000)). In other situations, however, the accuracy of background extra-statistical knowledge and the lack of reliable statistical information may suggest that we need to make extensive use of background knowledge.

As mentioned in Section 2.3, causal models rely on a hypothetico-deductive methodology. This view offers also a natural middle way between deductivist and inductivist approaches, which has to be established in practice in each real-world case.

4. Causal modelling and evidence

As mentioned in the introduction, in this paper we tackle the question of the *causal* interpretation of statistical models, and we frame the discussion specifically in the case of econometrics. The arguments given above hinge upon methodological considerations about the kind of assumptions, the amount of background knowledge needed and the methodology used in order to make the step from an associational model towards a causal model. We argued that, in a statistical model, we infer *associations* (or dependencies) between variables, while a causal model is an 'augmented' statistical model, 'augmented' in such a way that we can infer *causal relations*. In this section, we discuss that part of the augmentation having to do with the evidence generated in the associational model and in the causal model, notably evidence of difference-making and mechanisms. We try to build a bridge between the recent literature in the philosophy of medicine – where 'evidential pluralism' is currently being debated – and causal modelling in econometrics.

4.1 Evidential pluralism

The remote origins of evidential pluralism lie in the philosophical literature on causation developed in the second half of the twentieth century. The accounts developed in that period clustered around two main ideas.

On the one hand, in 'difference-making accounts', the claim '*A* causes *B*' is established in virtue of there being some appropriate difference-making relation (probabilistic, counterfactual or manipulation) between *A* and *B* (see, for instance, Arntzenius, 2010; Eells, 1991; Lewis, [1973] 1986, 2004; Reichenbach, 1956; Suppes, 1970). For instance, the causal claim 'Smoking causes lung cancer' is established, depending on the particular difference-making account embraced: (1) if smoking increases the chances of developing lung cancer or (2) if, had smoking rates decreased, cancer rates would have decreased too or (3) if intervening on smoking behaviour would result in decreased cancer rates.

On the other hand, in 'mechanistic accounts' the claim '*A* causes *B*' means that there is a mechanism, or a process, linking the cause to the effect (see, for instance, Bechtel & Abrahamsen, 2005; Craver, 2007; Dowe, 2000; Glennan, 2002, 2010; Machamer, Darden, & Craver, 2000; Salmon, 1984). For instance, the causal claim 'Smoking causes lung

cancer' is established to the extent that there is a sufficiently well understood mechanism that links smoking to lung cancer.

In a recent paper, Russo and Williamson (2007) develop an argument to show that there is a share of problems in holding one, the other or some combination of the above-mentioned views. The mistake, in each of the above-mentioned accounts, is to reduce the concept of causation to either the concept of mechanism or to (an appropriate form of) difference-making. Russo and Williamson (2007) go on in arguing that the main problem is the confusion between the very concept of causation and the *evidence* needed in order to establish a causal claim. Causal claims, in their view, are established on the basis of evidence of mechanisms *and* difference-making. They illustrate this thesis about evidential pluralism in the field of biomedical research, appealing to current practices in, e.g., the International Agency for Research on Cancer and to several examples from history of medicine.[9]

This form of evidential pluralism, now referred to as the 'Russo–Williamson Thesis' (RWT), generated a substantial body of literature in the context of medicine (see, for instance, Clarke, Gillies, Illari, Russo, & Williamson, 2013; Illari, 2011; La Caze, 2011; Russo & Williamson, 2011; Weber, 2009), but much less in the context of the social sciences and in particular econometrics. In the following, we discuss how different *evidential components* are generated in the associational and causal models presented in the previous sections and how they support the inferences drawn in associational and causal models.

4.2 Types of evidence and types of model

According to RWT, causal claims need support from two types of evidence, broadly conceived. One type is evidence of difference-making, that is probabilistic, statistical or counterfactual relations between variables of interest indicating that the putative causal factor is in fact relevant (or, makes a difference) to the occurrence of the putative effect. Difference-making evidence is generated, in a first instance, by the statistical information – in the form of dependencies – of associational models (see earlier Section 3.1). Difference-making is needed to justify the choice of building the 'augmented' statistical model discussed earlier in Sections 2 and 3. Difference-making evidence also supports claims about prediction and control derived from the causal claims established in the model.

The other type of evidence is evidence of mechanisms, that is knowledge – whether plausible or confirmed – about mechanisms that (may) underlie the observed dependencies. The issue of whether there are socio-economic mechanisms as opposed to individual-level mechanisms is not easy to settle (for a discussion see Little, 2006). An important part of the recent debates in social science methodology has been focussed on the characterisation and modelling of socio-economic mechanisms (see, for instance, Demeulenaere, 2011). In this literature, socio-economic mechanisms are not just the statistical aggregate (e.g. sum or average) of individual mechanisms, but may also display properties not reducible to the micro-level.

According to the econometric tradition which can be traced back to the Cowles commission, (economic) *theory* provides the mechanisms, and empirical research is eventually about testing the theory against available data (cf. Boumans, 2010). But we can also conceive of causal modelling as the modelling enterprise that we use to hypothesise and test (economic or social) mechanisms (see, for instance, Mouchart & Russo, 2011; Mouchart, Russo, & Wunsch, 2010; Russo, 2011b). In this view, the 'augmented'

statistical model codifies the available background knowledge and preliminary analysis of data in the 'recursive decomposition'. The recursive decomposition (a generalisation of the marginal-conditional decomposition mentioned in Section 2.2) breaks down the initial joint probability distribution into a sequence of marginal and conditional components that can be interpreted as a mechanism. The important feature of causal mechanisms, in this perspective, is that they carry explanatory power.

What RWT says is that, alone, neither evidence of difference-making nor evidence of mechanisms suffices to establish causal relations. This is typically the case in econometrics too, as we seek to establish dependencies (in associational models), *and* we also impose further constraints (the augmentation) in order to draw causal inferences. Mechanisms play an important role in the augmentation phase. This, notice, does not amount to saying that difference-making and mechanisms are necessary (or sufficient) conditions for causal relations. The thesis has epistemological and methodological scope. This means that RWT indicates what to look for in order to establish causal relations, i.e. what kind of evidence licences claims about dependency and what kind of evidence licences claims about causality.

Differently put, RWT is not a tick-list for items of evidence such that, if we have them, we are automatically allowed to interpret the model causally. Evidential pluralism refers to the way evidence generated in a model has to be evaluated. What we seek to evaluate is whether statistical and causal information allow us to establish that the putative causal factors make a difference to the effect, and that there is some mechanism that supports such difference-making relations. Thus, one single item can be evidence of both difference-making and mechanism. Examples from medicine are perhaps clearer in this respect. For instance, observations about the mode of transmission of a bacterium, say *Vibrio Cholerae*, provide evidence about the mechanism underlying such transmission *and* that the bacterium makes a difference to the occurrence of the disease.

In socio-economic contexts we are most often provided with evidence of difference-making generated by associational models, and the real challenge is to use or model causal information such that evidence of mechanisms supports causal inferences. One difficulty is that in socio-economic contexts, very different mechanisms can support the same difference-making relations. Population heterogeneity is a real problem for social research and it is an empirical question how much (and to what extent) the same mechanisms are at work in different contexts. Mechanisms are often highly context-dependent, but this does not mean they are of no use. Quite to the contrary, it is of utmost importance, when building the causal model, to be explicit about the assumptions and background knowledge employed during the various phases of building, testing and interpreting a model. This, as we shall discuss later, is crucial in conflict resolution, that is, when we have radically different explanations for the same phenomenon.

4.3 Some consequences of evidential pluralism

In this section we show that evidential pluralism is able to shed new light on three broad issues concerning causal modelling: model-dependence, fallibility and conflict resolution.

4.3.1 Model-dependence

A main consequence of evidential pluralism is that the validity of causal claims is relative to their respective models. This does not amount to denying the reality of socio-economic phenomena. It is instead a way of highlighting the importance of each stage and part of the

modelling enterprise: assumptions, statistical and causal information, background knowledge, data analysis and testing. In other words, results of empirical studies are primarily valid *within* the model being developed, and each stage of the modelling procedure contributes to establishing the validity of the model.[10]

We mentioned in Section 2 studies on household consumption. A well-established empirical regularity is the so-called *Engel's law*: the poorer is a family, the larger its budget share spent on nourishment (cf. Chai & Moneta, 2010, and references therein). This is indeed a well-known empirical generalisation, grounded on the numerous investigations which have applied statistical modelling to a wide range of data-sets and have found this form of statistical dependence between food budget share and income. As mentioned, the standard approach (in neoclassical economics) is to causally augment this statistical information and to embed it in a theoretical model which contemplates utility maximisation and rational choice. Other, more *naturalistic*, approaches provide alternative interpretations of that statistical dependence by pointing out at the hierarchical structure of individual needs, physiological wants, such as those motivating food consumption, have to be (to a certain degree) satisfied before devoting resources to consumption activities aimed at the satisfaction of other (e.g. psychological or cognitive) needs (Chai & Moneta, 2012; Witt, 2001).

However, such claims will be valid within the corresponding model being developed, namely depending on the analysed data, on the background knowledge of the researchers and on the methods used. It is an *empirical* question whether the same results are valid for different data-sets (namely for different populations) and whether different modelling strategies lead to different results, in which case it is of utmost importance to find out why.[11] Regarding food consumption, for example, it is interesting to note that Banerjee and Duflo (2011) found out that in Morocco families that do not have enough to eat buy a television and they analysed this phenomenon on the basis of an adequate and evidence-based understanding of poverty.

In summary, the 'augmentation' of statistical evidence (associations) in causal terms is an interpretation of real aspects of the reality against the backdrop of our background knowledge. Often, background knowledge is formalised (as in neoclassical economics) in terms of an analytical model, sometimes articulated in substantial information related to different scientific areas. This is expressed in a language familiar to econometricians and basically rephrases what we expressed earlier, in a language that is more familiar to philosophers: there is an interplay and mutual support of evidence of difference-making (the dependencies) and mechanisms (e.g. economic behaviour).

But causal claims in social research can be model-dependent in another, less direct, way. Statistical, or difference-making, evidence, on which associational models and causal models are based, is relative to the statistical model because it is generated within this frame. This simple fact renders causal claims (when inferred in the framework of an augmented statistical model) model-dependent. The fact that statistical evidence is relative to (or makes senses in) the statistical model where it has been generated has sometimes been neglected. In fact, data are often viewed as if they were already statistical evidence *per se*. But this view is not correct. Data are but a sample of an unobserved or hypothetical DGP, so it cannot be 'evidence'. Data *analysis*, in an associational model, provides evidence of difference-making. This neglect is apparent in empirical works (in economics or social research more generally) in which correlations or linear regression functions, which involve precise assumptions about the DGP, are estimated *before* testing any feature of the DGP. The consequence is that one may obtain bad sample estimates of population values parameters. This point has been clearly made already by Spanos (1990) and also by Hoover (2009).

4.3.2 Fallibility

We talk about fallibility in two senses. The first is that the *results of empirical studies* should be always considered fallible and open to revision, once a more reliable background knowledge or new data becomes available, or once different modelling strategies are tried out. Evidence is neither monolithic nor immutable, and we have to allow for the possibility of revising the results of studies on the basis of new evidence generated. We said earlier that background knowledge plays a crucial role in the 'augmentation' of statistical models. But this does not mean that background knowledge cannot be changed. There is instead a *va et vient* between established knowledge and knowledge being established, and the evidence generated in empirical studies of course looms large in this dynamic.

The second is that the *statistical evidence* (i.e. difference-making evidence encoded in the dependencies of the associational model) upon which causal claims are based should also be considered fallible and open to revision. This is related to the earlier point about model-dependence. It is always conceivable that a more appropriate formalisation of the DGP (perhaps due to progress in statistical methods or data collection) delivers a more accurate statistical model and hence more *reliable* statistical evidence.

The acknowledgement of fallibility reinforces the point about model-dependence in the sense that, precisely because we can get things wrong, it is of utmost importance to specify how we got to those results. However, this is not to undermine the possibility of drawing causal conclusions from models in social research.

4.3.3 Conflict resolution

The views about evidential pluralism and model-dependence, we maintain, also help with conflict resolution. A major gain in adopting the framework we suggested – notably, the account of associational versus causal models, evidential pluralism and model-dependence – allows a better and more fruitful analysis of conflict resolutions, i.e. when scientists disagree about causal relations. Disagreements may in fact come from different ways of specifying the model (that is different difference-making evidence) or from different pieces of information in the augmentation (that is different mechanisms to support dependencies). Thus, specifying how difference-making evidence has been generated or what (socio-economic) mechanisms have been called to support dependencies may help understand the origin of divergent results. This is not guarantee that a solution will be found, but at least helps to make clear where the dispute lies.

Consider, for instance, the dispute in the 1980s between Milton Friedman and Anna Schwartz, on the one hand, and David Hendry and Neil Ericsson, on the other hand, on money demand in the UK (Friedman & Schwartz, 1982, 1991; Hendry & Ericsson, 1985, 1991)[12]. In their 1982 book, *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*, Friedman and Schwartz present a number of empirical findings which support several economic hypotheses consistent with the quantity theory of money, according to which the money stock is the predominant factor explaining changes in money income. Hendry and Ericsson (1985) examined their empirical claims about monetary behaviour in the UK and found that these were established upon an econometric model which was not adequately specified.

In the regressions that Friedman and Schwartz (1982) estimated to establish evidence concerning the money demand, and the influence of money on income and prices in UK, they used phase-averaged data, considered the parameters in the money demand equation

as constant, treated money as exogenous and, more in general, ignored the mutual interdependence of money, income and prices, and interest rates. Hendry and Ericsson (1985) showed that these assumptions can be tested from the data (or at least have testable implications). Using the same UK data used by Friedman and Schwartz, Hendry and Ericsson rejected those assumptions, removing credibility from the theoretical conclusions of Friedman and Schwartz (1982). Moreover, they estimated a better-fitting, constant, dynamic error-correction (i.e. taking into account co-integrating relationships) model. This model turns out to be a model of money but not of prices because 'its constancy holds only conditionally on contemporaneous prices' (Hendry & Ericsson, 1991).

Two remarks are in order. First, questioning the results means to discuss different aspects of the model building and model testing process (methodology, assumptions, background knowledge). To ground both the claim that Friedman and Schwartz's model was mis-specified and the proposal of an alternative empirical model of money demand, Hendry and Ericsson (1991) discuss at length the methodology of model specification. In a manner consistent with Spanos' (and ours, see Section 2.2) perspective, Hendry and Ericsson (1991) see the statistical model as a DGP 'characterized by the joint density of the observable variables'. Empirical (econometric) models are just *transformation* or 'reduction' of the DGP and each step of reduction should be tested vis-à-vis the data (Cook & Hendry, 1994). For example, one of the implicit (or hidden) assumptions in Friedman and Schwartz's analysis is that disturbances are normal. Using a standard test, Hendry and Ericsson (1985) show that normality should be rejected. Friedman and Schwartz, on the other hand, use statistical analysis to find associations which have to be heavily interpreted by a priori historical knowledge about institutional arrangements and mechanisms (cf. Friedman & Schwartz, 1991).

The second remark we want to make is about fallibilism: in the light of new methods of analysis (or data collection) we can revise established results. If econometrics, since the time Hendry and Ericsson wrote, offers new tools, then it is a sensible thing to rerun the analysis and check whether the same results obtain. In a recent paper, indeed, Ericsson (2011) re-evaluates Hendry and Ericsson's (1985, 1991) model with *Autometrics*, which is an algorithm for computer-automated model selection (cf. Doornik, 2009), finding ground for several (minor) improvements.

It is worth noting, however, that this does not licence an 'anything goes' line of argument. If background knowledge is used in different manners and if there are competitive specifications of the statistical model, one should always question which alternative is better justified. These questions are not settled a priori. They are instead settled by an open and honest discussion about the choices made at each step of the modelling procedure. So we are in the position, at least in principle, to settle conflicts. It is also worth noting that this is no guarantee that conflicts will be resolved, as we may not reach a consensus about what background or testing technique to use. But surely open and honest confrontation at the level of the *model* should lead the debate forward, and thereby enhance our understanding of phenomena.

5. Conclusion

Econometrics, and the social sciences more generally, makes extensive use of statistical models in order to gain causal knowledge of phenomena. While early statisticians and methodologists such as Wright (in the 1930s) or Blalock (in the 1960s) were more prone to adopt a causal interpretation of statistical models, present-day scientists (and philosophers of science alike) are cautious in drawing causal conclusions from statistics. We also share

this cautious attitude, and yet we think that the question of when a statistical model can be given a causal interpretation needs to be addressed further.

The paper advances the view that we are justified in interpreting a statistical model in causal terms when it is 'augmented' with some specific features. We developed two lines of arguments. The first draws a distinction between associational models and causal models (and thereby between statistical and causal information); the second examines evidential pluralism in the context of econometric modelling.

Associational models are statistical models that allow us to infer dependencies between variables of interest. They are based on a minor use of background knowledge and assumptions. In order to infer *causal* relations, we need to augment associational models with a lot more background knowledge and assumptions, and to employ a hypothetico-deductive methodology. An important difference between associational models and causal models is that they encode, respectively, statistical and causal information.

Both statistical and causal information describe some facets of the DGP which realises the values of certain variables of interest. At bottom, the difference between statistical and causal information is that the former consists of symmetric relations (*X* is associated with *Y* implies that *Y* is associated with *X*), whereas the latter consists of asymmetric relations (*X* causes *Y* does not imply that *Y* causes *X)*. While statistical information describes how these values are regularly associated, or how the values regularly co-vary, causal information contributes to a description of the DGP as a 'mechanism'. Statistical and causal information thus contribute to making successful inferences about intervention, explanation and prediction.

We addressed the question of why causal information does not necessarily introduce a vicious circle. Causal information is statistical information augmented with background knowledge, namely assumptions about some facets of the DGP which also include causal propositions. Vicious circles are excluded if assumptions about causal propositions we want to validate are deliberately not included in this augmentation.

We then considered the recent literature on evidential pluralism, namely the thesis that causal claims are established on the basis of a plurality of evidential sources. The thesis has not received close attention in social contexts, particularly in econometrics. We argued that associational models generate evidence of difference-making, which is encoded in the dependencies inferred within such framework. Causal models also generate evidence of mechanisms, which is encoded in the formalisation of the DGPs.

Ultimately, we hold the view that a causal interpretation of statistical models is justified to the extent that statistical information is augmented into causal information, or in other words that evidence of difference-making is integrated with evidence of mechanisms.

We discussed three consequences of adopting such a view: model-dependence, fallibility and conflict resolution. These are all consequences of highlighting the importance of the modelling procedure, and of each step therein, for establishing causal relations. However, model-dependence does not imply that no generalisation about economic (or social) phenomena can be attained. It just says that it is an *empirical* question whether, and to what extent, we can generalise. Similarly, reminding about the fallibility of studies is not to undermine causal analysis at its very basis. It just says that our knowledge and understanding of economic (or social) phenomena is not immutable, and that it is precisely through further empirical studies that we can improve it. But this may mean to discard previous results. Finally, conflict resolution is a delicate issue, and we wanted to suggest that honest and open discussions about any stage of the modelling

procedure (from data collection to testing and interpretation) may shed light on divergent conclusions.

Acknowledgements

We are very grateful to the participants of the workshop 'The role of mathematical modelling in the social sciences' (A Coruña, September 2010), as well as the participants of the International Network for Economic Method (INEM) conference (Helsinki, September 2011), where we presented preliminary versions of this paper. We thank in particularly Marcel Boumans, David Hendry, Donald Gillies, Kevin Hoover, Ladislav Kvasz, Aris Spanos, two anonymous referees and the editors of this special issue for stimulating and helpful comments. Any error or imprecision remain of course ours. In the course of time, F. Russo has benefited from support given the British Academy (2009–2011) and from the Fonds Wetenschappelijk Onderzoek (FWO, Belgium) as Pegasus-Marie Curie Fellow (2012–2013).

Notes

1. On the interpretation of associational models and of causal models in terms of 'variation' rather than 'regularity', see Russo (2009b).
2. Since X_i, Y_i, W_i are i.i.d., we omit for convenience the subscript i. Notice that in this case $f(X_1, \ldots, X_n; \phi) = \prod_{i=1}^{n} f(X_i; \theta)$.
3. We thank one anonymous referee for suggesting this example.
4. The correlation is indeed a measure of dependence between two normal variates, and is defined in the following way: $\rho(X, Y) = \sigma_{XY}/\sigma_X, \sigma_Y$, where σ_{XY} denotes the covariance between X and Y; and σ_X, σ_Y are standard deviations. The partial correlation can be recursively defined as: $\rho(X, Y|W) = (\rho(X, Y) - \rho(X, W)\rho(Y, W))/(\sqrt{1 - \rho^2(X, W)}\sqrt{1 - \rho^2(Y, W)})$. Statistical dependence, denoted by $\perp\!\!\!\perp$ (Dawid, 1979), is defined as: $X \perp\!\!\!\perp Y$ iff $f_{XY}(x, y) = f_X(x) f_Y(y)$. Conditioning on W : $X \perp\!\!\!\perp Y|W$ iff $f_{XY|W}(x, y|w) = f_{X|W}(x|w) f_{Y|W}(y|w)$. More details are given in Section 3.1.
5. Or, rephrased in 'variation' terms alluded above, stochastic co-variations that are regular enough in the data-set.
6. The stochastic characteristic of the social phenomena is related to our incomplete *knowledge* of the world. This does not force us to take a stance about the metaphysical issue of whether the *world* is deterministic or indeterministic. In the framework hereby developed, we can meaningfully talk of stochastic social mechanism without any commitment to indeterminism.
7. In this way the 'black box' of the DGP is only partially opened.
8. That is, the causal Markov and the faithfulness condition (Pearl, 2000; Spirtes et al., 2000).
9. Russo and Williamson (2007) argue, together with the thesis about evidential pluralism, in favour of the 'epistemic theory', which is supposed to provide the causal metaphysics. Simply put, according to the epistemic theory, causation has to do with the (causal) inferences we are allowed to make on the basis of the available evidence, background knowledge and methods. Causation, in this theory, is primarily an epistemic category that we, as epistemic agents, use to chart the world in order to make successful inferences. The epistemic theory has also sparked debate, but we do not take stance with respect to this discussion here. Evidential pluralism is in fact compatible with other pluralistic, metaphysical views.
10. On this point, see also Russo (2011a).
11. Some readers may prefer the expression 'context-dependent' on the grounds that it is more general in scope than 'model-dependent'. We do not have any objection to that. In our approach, the context is part of the model and consequently there is not any incompatibility between the two.
12. For a discussion see Hammond (1996) and references therein.

References

Arntzenius, F. (Fall 2010). Reichenbach's common cause principle. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2010/entries/physics-Rpcc/

Banerjee, A., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. New York: Public Affairs.
Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and Statistics*, *79*, 527–539.
Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, *36*, 421–441.
Blalock, H. M. (1961). *Causal inferences in nonexperimental research*. Chapel Hill, NC: The University of North Carolina Press.
Boumans, M. (2010). The problem of passive observation. *History of Political Economy*, *42*, 75–110.
Chai, A., & Moneta, A. (2010). Retrospectives: Engel curves. *Journal of Economic Perspectives*, *24*, 225–240.
Chai, A., & Moneta, A. (2012). Back to Engel? Some evidence for the hierarchy of needs. *Journal of Evolutionary Economics*, 1–28.
Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*. doi:10.1016/j.ypmed.2012.10.020
Cook, S., & Hendry, D. F. (1994). The theory of reduction in econometrics. *Poznan Studies in the Philosophy of the Sciences and Humanities*, *38*, 71–100.
Craver, C. F. (2007). *Explaining the brain. Mechanisms and the mosaic of neuroscience*. Oxford: Oxford University Press.
Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*(1), 1–31.
Deaton, A. (1986). Demand analysis. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (vol. III, pp. 1767–1839). Amsterdam: Elsevier.
Demeulenaere, P. (Ed.). (2011). *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press.
Doornik, J. (2009). Autometrics. In J. Castle & N. Shephard (Eds.), *The methodology and practice of econometrics: A Festschrift in honour of David F. Hendry* (pp. 88–121). Oxford: Oxford University Press.
Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.
Dufour, J.-M., & Hsiao, C. (2008). Identification. In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). Basingstoke: Palgrave Macmillan. doi:10.1057/9780230226203.0762
Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.
Durkheim, E. ([1897] 1960). *Le suicide*. Paris: Presses Universitaires de France.
Eells, E. (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.
Ericsson, N. R. (2011). Empirical model selection: Friedman and Schwartz revisited. Unpublished manuscript. *European Economic Association Meeting*.
Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*, 309–368.
Friedman, M., & Schwartz, A. J. (1982). *Monetary trends in the United States and United Kingdom: Their relation to income, prices, and interest rates, 1867–1975*. Chicago: University of Chicago Press.
Friedman, M., & Schwartz, A. J. (1991). Alternative approaches to analyzing economic data. *The American Economic Review*, 39–49.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science. Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part II: Symposia Papers, September*, *69*, S342–S353.
Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research*, *81*, 362–381.
Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*(Supplement), 1–115.
Hacking, I. (1965). *Logic of statistical inference*. Cambridge: Cambridge University Press.
Hammond, J. (1996). *Theory and measurement: Causality issues in Milton Friedman's monetary economics*. Cambridge: Cambridge University Press.
Heckman, J. (2008). Econometric causality. *International Statistical Review*, *76*(1), 1–27.

Hendry, D. F., & Ericsson, N. R. (1985). Assertion without empirical basis: An econometric appraisal of Monetary trends in … the United Kingdom by Milton Friedman and Anna Schwartz. *International Finance Discussion Papers,* Number 270.

Hendry, D. F., & Ericsson, N. R. (1991). An econometric analysis of UK money demand in monetary trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz. *The American Economic Review*, *81*, 8–38.

Hoover, K. (2001). *Causality in macroeconomics*. New York: Cambridge University Press.

Hoover, K. (2009). Probability and structure in econometric models. *The Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science* (pp. 497–513). London: King's College.

Hoover, K. D., Johansen, S., & Juselius, K. (2008). Allowing the data to speak freely: The macroeconometrics of the cointegrated vector autoregression. *The American Economic Review*, *98*, 251–255.

Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, *25*, 139–157.

La Caze, A. (2011). The role of basic science in evidence-based medicine. *Biology and Philosophy*, *26*, 81–98.

Lewis, D. K. ([1973] 1986). Causation. *Philosophical papers*] (Vol. 2, pp. 159–213). Oxford: Oxford University Press.

Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 75–106). Cambridge, MA: The MIT Press.

Lindley, D. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*, 293–337.

Little, D. (2006). Levels of the social. In M. Risjord & S. Turner (Eds.), *The philosophy of anthropology and sociology* (pp. 343–371). New York: Elsevier.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

Moneta, A. (2007). Mediating between causes and probability: The use of graphical models in econometrics. In F. Russo & J. Williamson (Eds.), *Causality and probability in the sciences* (pp. 109–130). London: College Publications.

Mouchart, M., & Russo, F. (2011). Causal explanation: Recursive decompositions and mechanisms. In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 317–337). Oxford: Oxford University Press.

Mouchart, M., Russo, F., & Wunsch, G. (2010). Inferring causal relations by modelling structures. *Statistica*, *LXX*, 411–432.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Quetelet, A. (1869). *Physique sociale. Ou essai sur le développement des facultés de lhomme*. Bruxelles: Muquardt.

Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.

Russo, F. (2009a). Causal arrows in econometric models. *Humana. Mente*, *10*, 25–39.

Russo, F. (2009b). *Causality and causal modelling in the social sciences. Measuring variations.* Methodos Series. New York: Springer.

Russo, F. (2011a). Correlational data, causal hypotheses, and validity. *Journal for General Philosophy of Science*, *42*, 85–107.

Russo, F. (2011b). Explaining causal modelling. Or, what a causal model ought to explain. In C. Sinigaglia, & D. D'Agostino (Eds.), *Prooceedings of the SILF Conference 2007*.

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in Philosophy of Science*, *21*, 157–170.

Russo, F., & Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences*, *33*, 563–582.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Sober, E. (2001). Venetian sea levels, British bread prices, and the principle of the common cause. *The British Journal for the Philosophy of Science*, *52*, 331–346.

Spanos, A. (1990). The simultaneous-equations model revisited. Statistical adequacy and identification. *Journal of Econometrics*, *44*, 87–105.

Spanos, A. (1999). *Probability theory and statistical inference: Econometric modeling with observational data*. Cambridge: Cambridge University Press.
Spanos, A. (2005). Structural equation modeling, causal inference and statistical adequacy. In P. Hajek, L. Valdes-Villanueva, & D.W. (Eds.), *Logic, methodology and philosophy of science* (pp. 639–661). London: King's College.
Spanos, A. (2006). Where do statistical models come from? Revisiting the problem of specification. *Lecture Notes-Monograph Series*, *49*, 98–119.
Spanos, A. (2007). Curve fitting, the reliability of inductive inference, and the error-statistical approach. *Philosophy of Science*, *74*, 1046–1066.
Spanos, A. (2010). Graphical causal modeling and error statistics. Exchanges with Clark Glymour. In D. G. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability and the objectivity and rationality of science* (pp. 364–375). Cambridge: Cambridge University Press.
Spanos, A. (2011). Revisiting the statistical foundations of panel data modeling. Unpublished manuscript.
Spanos, A. (2012). Learning from data: The role of error in statistical modeling and inference. Unpublished manuscript.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: The MIT Press.
Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland Publishing.
Weber, E. (2009). How probabilistic causation can account for the use of mechanistic evidence. *International Studies in Philosophy of Science*, *23*, 277–295.
Witt, U. (2001). Learning to consume – A theory of wants and the growth of demand. *Journal of Evolutionary Economics*, *11*, 23–36.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557–585.